Validation of Canopy Height Profile methodology for small-footprint full-waveform airborne LiDAR data in a discontinuous canopy environment

ISPRS Journal of Photogrammetry and Remote Sensing 104 (2015) 144–157

Contents lists available at ScienceDirect

ISPRS Journal of Photogrammetry and Remote Sensing

journal homepage: www.elsevier.com/locate/isprsjprs

Validation of Canopy Height Profile methodology for small-footprint full-waveform airborne LiDAR data in a discontinuous canopy environment

Karolina D. Fieber [a,*], Ian J. Davenport [b], Mihai A. Tanase [c], James M. Ferryman [d], Robert J. Gurney [a], Victor M. Becerra [d], Jeffrey P. Walker [e], Jorg M. Hacker [f]

[a] University of Reading, School of Mathematical and Physical Sciences, Harry Pitt Building, Reading RG6 6AL, UK
[b] University of Reading, School of Archaeology, Geography and Environmental Sciences, Whiteknights RG6 6DW, UK
[c] University of Melbourne, Parkville, Melbourne, 3010 VIC, Australia
[d] University of Reading, School of Systems Engineering, Whiteknights RG6 6AY, UK
[e] Monash University, Faculty of Engineering, Clayton, Melbourne, VIC 3800, Australia
[f] Airborne Research Australia, School of the Environment, Flinders University, Adelaide, SA 5001, Australia

ARTICLE INFO

Article history:
Received 12 November 2014
Received in revised form 23 January 2015
Accepted 5 March 2015

Keywords:
Full-waveform
LiDAR
Small-footprint
Canopy height profile (CHP)
Effective leaf area index (LAIe)
Discontinuous canopy cover

ABSTRACT

A Canopy Height Profile (CHP) procedure presented in Harding et al. (2001) for large footprint LiDAR data was tested in a closed canopy environment as a way of extracting vertical foliage profiles from LiDAR raw-waveform. In this study, an adaptation of this method to small-footprint data has been shown, tested and validated in an Australian sparse canopy forest at plot- and site-level. Further, the methodology itself has been enhanced by implementing a dataset-adjusted reflectance ratio calculation according to Armston et al. (2013) in the processing chain, and tested against a fixed ratio of 0.5 estimated for the laser wavelength of 1550 nm. As a by-product of the methodology, effective leaf area index (LAIe) estimates were derived and compared to hemispherical photography values. To assess the influence of LiDAR aggregation area size on the estimates in a sparse canopy environment, LiDAR CHPs and LAIes were generated by aggregating waveforms to plot- and site-level footprints (plot/site-aggregated) as well as in 5 m grids (grid-processed). LiDAR profiles were then compared to field biomass profiles generated based on field tree measurements. The correlation between field and LiDAR profiles was very high, with a mean R^2 of 0.75 at plot-level and 0.86 at site-level for 55 plots and the corresponding 11 sites. Gridding had almost no impact on the correlation between LiDAR and field profiles (only marginally improvement), nor did the dataset-adjusted reflectance ratio. However, gridding and the dataset-adjusted reflectance ratio were found to improve the correlation between raw-waveform LiDAR and hemispherical photography LAIe estimates, yielding the highest correlations of 0.61 at plot-level and of 0.83 at site-level. This proved the validity of the approach and superiority of dataset-adjusted reflectance ratio of Armston et al. (2013) over a fixed ratio of 0.5 for LAIe estimation, as well as showed the adequacy of small-footprint LiDAR data for LAIe estimation in discontinuous canopy forests.

1. Introduction

Vegetation plays a very important role on Earth, as through its interaction with the atmosphere it is responsible for the exchange of energy fluxes (Breda, 2003; Levy and Jarvis, 1999). The information about its amount and spatial distribution is a key factor in many environmental studies and applications (Zhao et al., 2011) including flood modelling, fire risk assessment, carbon stock modelling (Lefsky et al., 2005b; Koetz et al., 2006) and global environmental change (Jonckheere et al., 2004). With the foliage distribution variable in horizontal and vertical directions as well as with time (seasonally and over years) (Breda, 2003), vegetation's complicated structure is very difficult to quantify and map in a timely manner. Field inventories are time consuming, laborious, expensive and not always accurate enough, often relying on allometric equations or other ground-based indirect methods while direct methods are highly destructive. Such field methods

* Corresponding author. Tel.: +44 (0) 118 378 5215; fax: +44 (0) 118 378 6413.
E-mail address: k.fieber@reading.ac.uk (K.D. Fieber).

http://dx.doi.org/10.1016/j.isprsjprs.2015.03.001
0924-2716/

are therefore unsuitable for long-term monitoring of large areas (Jonckheere et al., 2004) due to low frequency of sampling and high costs. Remote sensing methods on the other hand, both airborne and spaceborne, provide non-destructive and fast tools for obtaining better and wider coverage and therefore they are particularly useful for monitoring and inventories of vegetation at regional and global scales.

The two parameters commonly used to describe vegetation amount, derived from remote sensing observations are: normalized difference vegetation index (NDVI) and leaf area index (LAI). NDVI is a well-established passive optical remote sensing technique relying on the difference in vegetation's reflectance in the red and infrared parts of the electromagnetic spectrum. NDVI and other optical remote sensing techniques provide no information on the vertical distribution of the canopy, though, due to their two-dimensional character (no elevation information) (Morsdorf et al., 2006). They are therefore incapable of providing vertical foliage profiles. They are also sensitive to many factors such as e.g. atmospheric and soil effects, and their accuracy depends on the pixel size relation to the size of vegetation components (Riaño et al., 2004). This method has also been reported to saturate at high level vegetation biomass (Chen and Cihlar, 1996).

Leaf area index (LAI), defined as half the total leaf area per unit ground area (Lang et al., 1991; Chen and Black, 1992), has also been a widely studied parameter. Despite the simplicity of its definition, it is still one of the most difficult parameters to measure (Breda, 2003; Richardson et al., 2009). The methods of obtaining LAI are commonly divided into two groups: direct and indirect measurements (Breda, 2003; Jonckheere et al., 2004; Levy and Jarvis, 1999; Chen et al., 1997). Extensive reviews of ground-based direct and indirect methods of LAI retrievals with their advantages and disadvantages are presented in Jonckheere et al. (2004) and Breda (2003), with a background theory summary described in Weiss et al. (2004).

1.1. Leaf area index

Remote sensing methods of LAI estimation fall into the indirect measurements group, and as such are based on the Beer–Lambert law and a statistical approach to canopy element distribution within the crowns (Breda, 2003) that assumes that the foliage material arrangement within the canopy is random. All indirect remotely-sensed methods provide so-called effective LAI (LAIe) estimates. This term was first proposed by Black et al. (1991) and relates to the fact that in reality, the assumption of randomness of foliage is often violated. This is particularly the case for coniferous forests, where the clumping of canopy elements strongly affects the retrievals of LAI. Furthermore, remotely-sensed methods do not distinguish leaf area from other canopy elements such as stems or branches (Levy and Jarvis, 1999), thus providing plant area index (PAI) rather than LAI. Methods for correcting both for the clumping effect and for the contribution of woody elements of the canopy have been developed and are described in the literature (Chen et al., 1991; Chen et al., 1997; Breda, 2003). Following Jonckheere et al. (2004) who suggested the term effective LAI (LAIe) as the most intuitive for retrievals that have not been corrected for either of the above effects, in this study the LAIe term will be used.

In recent years, remotely-sensed LAI has commonly been derived from an active technique – LiDAR (Light Detection And Ranging). LiDAR, due to its ability to penetrate through the canopy gaps, offers the possibility to generate vertical profiles of three-dimensional vegetation structure. Therefore, not only the amount of vegetation (biomass) can be quantified but also the vertical distribution of foliage and its changes with the height above ground level can be described. Laser scanning has been tested for LAIe estimation by many scientists using a range of platforms (terrestrial, airborne and spaceborne). Most airborne small-footprint LiDAR studies have used the discrete point frequency ratios to predict leaf area or vegetation cover (Solberg et al., 2006; Solberg et al., 2009; Morsdorf et al., 2006; Riaño et al., 2004; White et al., 2000; Jensen et al., 2008) although some have also used statistical height metrics such as percentiles, variance, skewness, kurtosis etc. (Jensen et al., 2008). The LAIe estimates from discrete return data do however suffer from some shortcomings. By not accounting for intensity and treating all pulses equally regardless of how much light was returned and how much vegetation they therefore actually represent, they may lead to certain inaccuracies in the estimates (e.g. overestimation of ground returns and thus underestimation of LAIe). Relatively high thresholds used in the pulse detection methods can cause some of the returns to be missed and therefore bias the results, especially if vertical foliage profiles are sought. Subsequently, some of the weak ground returns may remain undetected, resulting in overestimation of low value LAIe and saturation of LAIe at high values, especially for small areas, when the Beer–Lambert theory is used (Richardson et al., 2009).

The last decade has seen a rapid development in the commercial small-footprint full-waveform laser scanner sector, making those instruments easily available and of widespread use. Such instruments offer additional benefits (e.g. denser vertical sampling, estimates of width, amplitude and backscatter cross-section of the pulses etc.) over point clouds generated by conventional discrete return laser systems (Mallet and Bretar, 2009). They may help overcome some of the drawbacks of discrete systems by utilising the raw light curve of the returned energy or allowing a custom decomposition procedure better suited to vegetation analysis. Several studies have investigated the usefulness of full-waveform LiDAR data for vegetation description using ground-based LiDAR systems (Hopkinson et al., 2013; Zhao et al., 2011; Zhao et al., 2012; Yao et al., 2011; Zheng and Moskal, 2012; Lovell et al., 2003; Hosoi et al., 2010), airborne platforms (Tang et al., 2012; Harding et al., 2001; Lindberg et al., 2012; Lefsky et al., 1999; Sun and Ranson, 2000; Lefsky et al., 2005b; Morsdorf et al., 2009; Armston et al., 2013; Calders et al., 2012; Hopkinson et al., 2013; Lovell et al., 2003; Hosoi et al., 2010; Ni-Meister et al., 2010) and spaceborne platforms (Miller et al., 2011; Lefsky et al., 2005a). Although ground-based systems provide very detailed information about the vegetation structure, their use, as in the case of any other ground method, is limited to small areas. They therefore provide an excellent source of information for calibration of the airborne and spaceborne LiDAR methods, which are more suitable for larger scale mapping.

With small-footprint waveform LiDAR data becoming increasingly popular, it seems natural to explore their potential. Small-footprint full-waveform LiDAR data, when used for LAIe estimation with the Beer–Lambert law, will suffer from similar saturation limitations as discrete data (Richardson et al., 2009), but possibly to a lesser extent. A custom decomposition procedure focused on detecting weak pulses will help reduce the possibility of missed ground returns, which is further minimized by using the raw-waveform rather than decomposed points. Nevertheless, as a result of the smaller beam cross-section (in comparison to large-footprint systems), not all of the small-footprint returned waveforms will include light which has reached the ground level, as some of them will be fully intercepted by vegetation. Therefore, aggregation into larger grids or discs is necessary. The main constraint dictating the size of grids for aggregated waveforms for LAIe calculation is the laser pulse density. The choice of grid size will also be a function of heterogeneity of the test area. For discontinuous or sparse canopies, large-footprint laser data (such as SLICER or LVIS) will most likely underestimate LAIe, whereas small-footprint data with

optimal grid size (adjusted to crown widths) may be able to capture in-between crown gaps and provide a more realistic estimate of LAI. Small-footprint full-waveform LiDAR may therefore be particularly suited for sparse canopies with a large variety of between-crown gaps.

1.2. Vertical foliage profiles

The vertical variation of foliage (or of leaf area) is often called foliage or canopy height profile (Aber, 1979; Lefsky et al., 1999; Harding et al., 2001; MacArthur and Horn, 1969). There are two groups of ground methods for foliage height profiles estimation: a destructive stratified clipping (Lefsky, 1997) and non-destructive point-quadrat sampling (Wilson, 1959; Wilson, 1965). Wilson's point-quadrat method employing a number of lines passing through the canopy and recording the points of intersection with leaves was modified by MacArthur and Horn (1969) to use an upward pointing camera with a range finding photo lens to determine the location of the first intersecting leaf (optical point-quadrat method). Assuming a random distribution of leaves these data at the lowest leaf level could then be transformed into vertical distribution. This procedure was further tested by Aber (1979) and was found to provide a foliage height profile consistent with those of standard point-quadrat method.

The MacArthur and Horn's (1969) foliage height profile methodology was subsequently adapted for use with LiDAR remote sensing as canopy height profiles (CHP) by Lefsky (1997) and further modified by other authors (Harding et al., 2001; Lefsky et al., 1999). Studies of vertical vegetation structure were mostly carried out with large-footprint full-waveform instruments. The algorithms were developed for NASA's airborne platforms such as SLICER (Harding et al., 2001; Lefsky et al., 1999) and LVIS instruments (Tang et al., 2012; Ni-Meister et al., 2010). Lefsky et al. (1999) presented a three-dimensional volume method of description of the canopy (Canopy Volume Method-CVM), summarizing the total volume and spatial organization of filled and empty space within the canopy. Harding et al.'s (2001) SLICER canopy height profile (CHP) procedure was found to reliably represent the vertical structure of a closed canopy and be closely related to the ground-based measurements in the stands tested in that study.

1.3. Related work and contributions

For small-footprint full-waveform airborne systems Lindberg et al. (2012) presented a methodology that used raw LiDAR waveforms to derive vegetation volume profiles and found the lowest RMSE and 'good correspondence' between field–derived and normalized waveform profiles, compared to three other methods tested. The study of Armston et al. (2013) focused on derivation of gap fraction and vegetation – ground reflectance ratio, which are the first step to LAIe estimation. However, in that particular study the waveforms were reconstructed from decomposition parameters due to lack of raw data, so the results are sensitive to the characteristics of the decomposition algorithm. A limited validation of this reconstruction with raw light curves was provided for a subset of data, proving good agreement. Chen et al. (2014) further tested the methodology of gap fraction estimation proposed by Armston et al. (2013) showing that the method is stable for different off-nadir scan angles, slopes of up to 26°, different acquisitions and different LiDAR systems. They also compared the derived LiDAR canopy gap fraction to gap fraction estimated from hemispherical photography achieving a Pearson correlation coefficient of 0.91.

Harding et al.'s (2001) SLICER CHP methodology was adapted to small-footprint LiDAR data by Fieber et al. (2013b) where preliminary results of LAIe validation (one of the stages of the procedure) against hemispherical photos in four sites were presented. The validation of CHP procedure from small-footprint data was also carried out at single-tree-level (Fieber et al., 2014) reaching R^2 of 0.86 for the LiDAR CHP of six combined swaths over a dead *Callitris glaucophylla* tree against field tree profile generated from convergent photographs. The raw waveform LAIe estimates in that study were found to be within ±5% of the hemispherical photography value.

This article is focused on validation of CHP methodology from small-footprint full-waveform LiDAR at plot- and site-level both for LAIe estimates as well as for vertical foliage profiles in a discontinuous canopy environment, which to the best of the author's knowledge has not yet been done. Further, the CHP procedure is enhanced by inclusion of the Armston et al. (2013) algorithm to determine dataset-adjusted vegetation-ground reflectance ratio. The results of CHP methodology with this ratio are then compared to the result of this same procedure using a fixed ratio of 0.5 (estimated for a laser wavelength of 1550 nm). Finally, the high resolution of the laser data offers the possibility to investigate LAIe and CHPs at different spatial scales, which is particularly important in a discontinuous canopy forest environment where the assumption of randomness of foliage distribution can be violated. In this paper, to test the impact of the size of aggregation area on LiDAR estimates, the data were aggregated into 5 m, plot-size (30 m diameter) and site-size (100 m diameter) cells.

The validation has been performed with the use of field tree measurements that have been converted into vertical field biomass profiles (for CHP) as well as based on hemispherical photography (for LAIe). Hemispherical photographs, also called hemi-photos (wide-angle field of view), provide a permanent record of the sky obstructions at the time of the shot. This method has been widely tested and used for the purpose of retrieving leaf area index (Chen et al., 1991; Macfarlane et al., 2007; Levy and Jarvis, 1999) and is now, with high resolution digital cameras easily available, a well-established technique for obtaining this index (Zhao et al., 2012). It has been previously used as a validation tool for other remote sensing methods, in particular for LiDAR LAIe retrievals (Zhao et al., 2012; Zhao et al., 2011; Richardson et al., 2009; Riaño et al., 2004; Morsdorf et al., 2006; Solberg et al., 2006) and so has been in this study.

2. Study area and data

The data used in this study were acquired as part of the Soil Moisture Active Passive Experiment 3 (SMAPEx-3) carried out in Australia in September 2011 (Panciera et al., 2013; Monerris et al., 2011). The study area, Gillenbah forest, is situated South of the town of Narrandera, between 447,548 m and 457,016 m (Easting) and 6,143,546 m and 6,149,810 m (Northing) (UTM, zone 55 H). The forest is about 3300 ha in area with White Cypress pine (*C. glaucophylla*) as the dominant species (90%), and the occasional Grey Box (*Eucalyptus microcarpa*) (10%). It is a forest with relatively discontinuous and varying canopy cover.

2.1. Field data

Twelve sites with different canopy cover were selected within the Gillenbah forest (Fig. 1). Each site consisted of five 500 m^2 plots located in the center and in each cardinal direction (Fig. 1, inset). All trees with a diameter larger than 5 cm were measured in the field, whereas the remaining trees were counted with their average height noted. The measurements included tree height (clinometer), crown base height (clinometer), circumference at breast height (tape), coordinates (GPS) and species. Upward-pointing hemispherical photographs were taken at each plot during the field

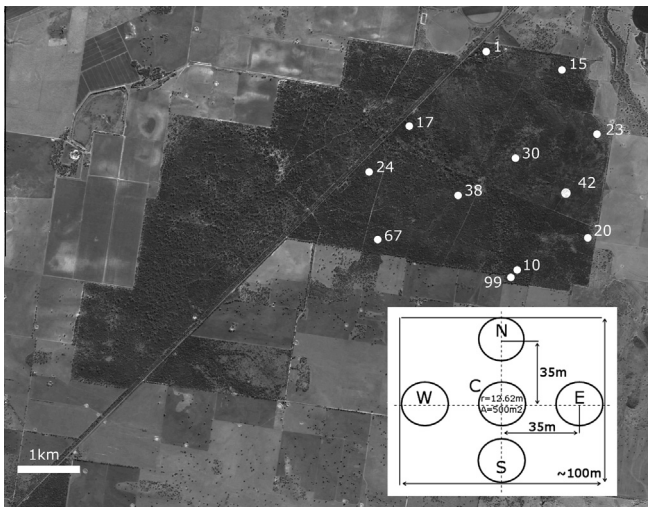

Fig. 1. Advanced Land Imager (ALI) image of Gillenbah forest study area with the twelve study sites locations overlaid. *Note*: Site markers are not to scale. Site design showing the layout of the plots is given in the inset (image: courtesy of Dr. Mihai Tanase).

survey, with a Nikon D40 camera mounted on a levelled tripod approximately 50–100 cm above ground level and a Sigma 4.5 mm fish-eye lens. Five hemispherical photographs, located at the center (C) and at each cardinal direction (N, E, S, W) were acquired at each plot. Although trees in twelve sites were measured in the field, the measurements in Site 1 were found to be incomplete. This site lacked cardinal direction hemispherical photographs as well as 70% of crown base height measurements. That in combination with the low number of trees and lack of understory made it impossible to generate a reliable and good quality field biomass profile or representative LAIe estimates for comparison with LiDAR values. Therefore Site 1 had to be excluded from the analysis and will not be discussed further.

2.2. LiDAR data

The full-waveform LiDAR data that were available for this study were acquired on 6th September 2011 with a Riegl LMS Q560 scanner. The laser instrument was mounted on a light aircraft and flown at 350 m above ground level (AGL) resulting in 0.18 m footprint diameter. The average laser shot spacing was 9 points/m^2 with an average detected point density of 19 points/m^2 (after decomposition). Both transmitted and received waveforms were recorded with a frequency of 1 GHz (1 ns spacing). The LiDAR data corresponding to field measurements at plot-level (within a cylinder of 15 m radius to cover the crowns of the trees at the edge of the plot) and site-level (within a cylinder of 50 m radius from the center of the central plot to cover the crowns of the trees at the edge of the site) were extracted from the swaths of data for each of the eleven sites. Despite large overlap between the LiDAR swaths which were acquired in a criss-cross pattern, only one swath per site was selected and used for the Canopy Height Profile (CHP) analysis. The extraction from the instrument was carried out using the GeoCodeWF commercial software and then the data were decomposed using custom Gaussian decomposition with a Trust-Region-Reflective optimisation algorithm (according to Fieber et al. (2013a)). Decomposed data of single-peak ground returns (identified with the help of backscattering coefficient) were subsequently used to generate Digital Terrain Models (DTM) for each of the sites.

2.3. Hemispherical photography LAIe

Upward-pointing hemispherical photographs (Fig. 2) were processed using HemiView software, a popular tool for estimating gap fraction and leaf area index (Riaño et al., 2004; Richardson et al., 2009; Solberg et al., 2006; Zhao et al., 2011). HemiView computes the fraction of sky obscured by vegetation in sky sectors, and turns it into leaf area estimates defined as half of the total leaf area per unit ground area. LAIe estimates were obtained for each of the five photographs taken at each plot. Plot-level estimates were, however, computed as the mean of four photographs taken at cardinal directions disregarding the central photograph. This was done due to considerable overlap between the central photograph and photographs at cardinal directions and due to the discontinuous character of the Gillenbah forest canopy. While in a dense forest adding the central photograph to the average would not have much impact on that average, because of discontinuous character of the site it would have somewhat biased the results toward the center of the plot. For the same reason, the central plot in site 38 had to be excluded from the plot level LAIe analysis, reducing the number of plots from 55 to 54. The central plot of site 38 had only one hemispherical photo taken in the field (no cardinal direction photos), which would not be representative of the whole plot. The site-level LAIe was computed as an average of the plot-level LAIe measurements in each of the eleven sites (Site 38 was included in the site-level analysis).

The clumping effect of canopy elements was neglected. Therefore the presented values of leaf area index represent a so-called effective leaf area index (LAIe). To obtain a value of true LAI, the LAIe would have to be multiplied by a clumping index specific to this study area and to the tree species. A final remark has to be made: the acquired photographs were not ideal, as being taken during the field work throughout the day, some of them had the sun present in the frame making the selection of threshold for LAIe calculation quite difficult and introducing some errors in the estimates. These photographs were nevertheless the only available for the measured sites.

2.4. Field biomass profiles

The tree parameter data collected during the field campaign were used to estimate leaf, branch, stem (biomass of tree trunk up to crown base height), stemwood (total tree stem biomass) as

Fig. 2. An example of a hemispherical photograph.

well as total above-ground biomass (in kg) for each measured tree within plot limits, and at plot- and site-level (in tons/ha). These estimates were calculated based on species-specific allometric equations (Burrows et al. (2001), Hamilton et al. (2005)) developed for a different study area. For *C. glaucophylla* two different allometric equations were considered: one based on diameter at breast height (DBH) and one based on height. The available equations for that species were developed for the South-central Queensland region (1000 km North of Gillenbah the study area) which is characterised by considerably higher annual average precipitation (600 mm). Since these models consistently overestimated tree heights for the Gillenbah forest study area when based on DBH, the biomass values obtained from DBH- and height-based allometric equations were averaged. Due to the lack of specific equations for *E. microcarpa*, the tree crowns being difficult to describe by a single geometric shape, and the small number of those trees (only 7.3% of measured trees) in the study site, the same equations were used for both species.

The biomass estimates for each tree were subsequently used to generate field vertical biomass profiles, with 15 cm bin vertical resolution at plot- and site-level, in order to facilitate the validation of LiDAR-based CHPs. The biomass was distributed across the tree height, weighted according to the volume of a composite truncated cone (Fig. 3). First, stem circumference at 0.3 m above ground level (x_a) was calculated from the Burrows et al. (2001) allometric equations based on measured circumference at 1.3 m (y_b): $y_b = -1.691 + 0.895x_a$[m]. Circumferences were then converted into diameters at 0.3 m (d(0.3 m)) and at 1.3 m (d(1.3 m) = DBH) and used to compute stem diameter at crown base height (d(CBH)) based on geometric proportions assuming a trapezium cross-section of the tree stem. If crown base height (CBH) was not measured in the field, breast height (1.3 m) was used in its place. Based on analysis of a few terrestrial photographs the stem diameter at the top of the tree was assumed to be 20% of DBH.

Crown base height and tree height were then assigned to their closest 0.15 m height bin. Knowing their diameters and the number of bins between them, the volumes of decreasing diameter truncated cones representing each 15 cm bin were calculated. They were then normalized by the total volume of the cones covering the crown region and used as weight to distribute leaf biomass across the crown depth. A similar procedure was used for stem biomass distribution across stem height, but using a cylinder (from 0 to 0.3 m) and truncated cone volumes calculated based on diameters at 0.3, DBH and CBH, respectively. For dead trees, stemwood biomass was distributed across the entire height of the tree summing the total volume along the tree height. Small (DBH < 5 cm) trees, were modelled in the same way as live measured trees but at plot-level rather than single-tree-level. Finally, the biomass of all the single trees modelled within each plot/site, and the corresponding plot/site-level model of small trees were aggregated in 15 cm bins to yield vertical profiles of field biomass.

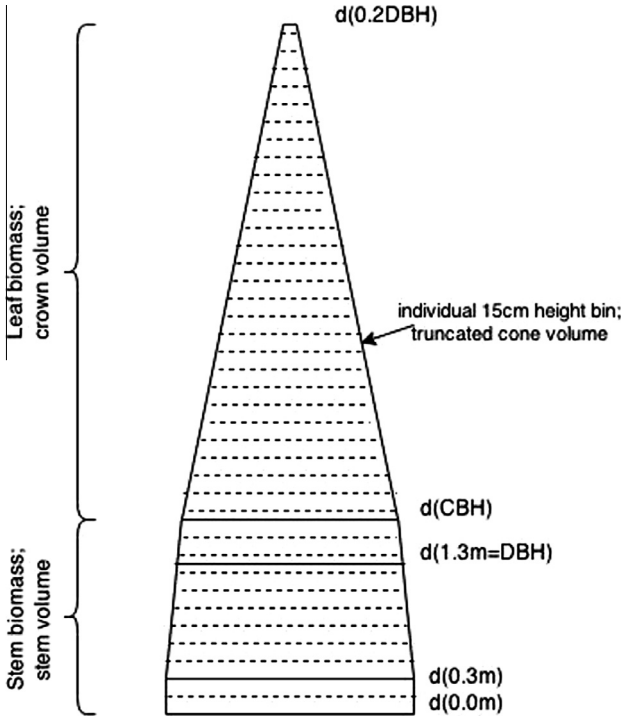

Fig. 3. Schematic diagram of weighting system applied to generate field biomass profiles.

3. Methods

The raw-waveform LiDAR methodology used in this study is based on the SLICER Canopy Height Profile processing presented in Harding et al. (2001). This methodology, originally applied to large-footprint data over a closed canopy environment, was adapted to small-footprint LiDAR and initially tested by Fieber et al. (2013b) at plot- and site-level in four forest sites for LAIe estimates (one of the stages of the procedure), in a comparative study of discrete point LAIe extraction methods. The technique was also tested for LAIe and vertical profile generation, however, only at single-tree-level (Fieber et al., 2014) with a very promising result (R^2 of 0.86 with field profile). Here, the validation of LAIe and vertical foliage profiles (CHPs) was performed at plot- and site-level using 11 field-measured sites in the discontinuous canopy cover of Gillenbah forest. The methodology was also enhanced by using the dataset-adjusted vegetation-ground reflectance ratio proposed by Armston et al. (2013) (WF1) and compared to the results of the procedure with a constant ratio of 0.5 (WF2) for the laser wavelength 1550 nm. The LAIe estimates were then compared to hemispherical photography derived values, and the vertical foliage profiles were compared to field biomass profiles.

To assess the quality of the results, ordinary least square regression analysis was carried out, and root mean squared error was calculated for each LiDAR LAIe comparison. Furthermore, a set of paired two-tailed t-tests (at the 5% significance level) of the null hypothesis that the differences between LiDAR and hemi-photo estimates were a random sample from a normal distribution with mean 0, against the alternative that the mean is not 0 were performed. Similarly to the LAIe analysis, for vertical foliage profiles a bin-wise ordinary least squares regression was undertaken, and root mean squared error calculated. Furthermore, a set of paired bin-wise two-tailed t-tests was carried out in order to assess whether the LiDAR profiles were significantly different from field-biomass profiles. T-tests were also performed between CHPs extracted using different vegetation ratios and between plot/site-aggregated and grid-processed CHPs.

3.1. Aggregation area

In comparison to large-footprint laser scanning data, small-footprint laser waveforms do not always have a ground return. Lack of a ground return in a waveform means that gap fraction cannot be estimated for it. Such a waveform therefore cannot be used in the processing on its own and has to be aggregated with others in a larger area. The selection of the aggregation cell size was thoroughly discussed in Richardson et al. (2009) and depends mostly on pulse density, sensitivity of the detection algorithm to weak returns, and site heterogeneity. For discontinuous canopies as in

the case of this study, aggregating data into plot- or site-size footprints was not ideal. When large between-canopy gaps are present in the aggregation cell, the leaf area will be computed for larger ground extents than the vegetation actually occupies. This will make the probability of pulses reaching the ground rise and in consequence, following the logarithmic transform, the LAIe of the examined area can be severely underestimated. The LAIe estimation based on gap fraction assumes the vegetation to be randomly distributed. In the case of the heavily clumped discontinuous canopies of this study area, aggregating the data into plot- or site-level cells will heavily violate the randomness assumption. Therefore, a cell of 5 m by 5 m was chosen at both plot- and site-level, to reduce the impact of clumped canopy on LAIe estimates, these are referred to as the grid-processed datasets.

Due to the circular nature of the plots and sites, the LAIe and CHP in every grid cell needed to be weighted proportionally to the area that the data covered in that cell – the cells on the edges of the plot/site were not fully covered by LiDAR (Fig. 4). A mean LAIe per plot/site was then calculated as a sum of weighted grid LAIe values in that plot/site divided by the sum of weights. In the rare situation of LAIe saturation within a grid cell due to a lack of ground returns, the maximum weighted grid LAIe value found in the particular plot/site was used as an approximation. That maximum grid LAIe value was then weighted by the area (weight) of the cell that has saturated and assigned to it. This was done to avoid removing saturated cells from the datasets and biasing the LAIe estimate. Another possible solution could have been used here. As proposed by Richardson et al. (2009) a number of ground returns within that cell could be artificially set to 1 for saturated cells. This procedure is easy to implement when LAIe is calculated from discrete returns. However, when waveform data are used, an artificial ground waveform would have to be used, which is not easily implemented due to energy variation between the laser shots. In the case of vertical foliage profiles, saturated cells had to be excluded from the profile generation (together with the weight assigned to the cell) in order to avoid distortion of the profile. The estimates of LAIe and CHP calculated in a 5 m by 5 m grid (referred to as grid-processed and denoted with letter 'g' e.g. WF1g) were compared to the estimates performed on the whole non-gridded datasets of plot and site radii (referred to as plot/site-aggregated), to see the impact of the aggregation area on LAIe and CHP retrieval from LiDAR data in discrete canopy environment.

3.2. Canopy Height Profile methodology

Canopy Height Profile (CHP) represents the relative vertical distribution of canopy surface area (vertical vegetation profile), and accounts for occlusion of the laser energy by the canopy. The CHP procedure is performed in several stages on raw-waveform data. The advantage of the waveform method over point methods (when points are sourced from full-waveform data) is that it does not require the full decomposition with optimisation procedure or calibration of the data. The knowledge about the approximate position of the first and the last pulse within the waveform obtained after the initialisation step of decomposition is sufficient, which is much less time-consuming and requires much less processing power than full decomposition. There is a need for DTM information though, but if the full decomposition is not performed, the DTM can also be derived from other sources of data or approximated from the initialisation step.

3.2.1. Waveform alignment

First, raw waveforms within the chosen aggregation area are aligned according to their elevation above ground level. This is carried out with the use of an existing DTM (generated based on single-peak ground returns) and the location of the first pulse. Due to the 1 ns sampling of the LiDAR system used, the vertical location of the first pulse in a waveform is associated with a 15 cm height bin. At this stage noise within the data is subtracted (estimate from the decomposition procedure or amplitude of the last waveform bin).

3.2.2. Returned energy profile

Once the waveforms are aligned, a returned energy profile is generated (Fig. 5A). The area underneath each waveform is calculated as a set of trapeziums: the average of each pair of consecutive amplitudes multiplied by the 0.15 m height change represented by each bin. The bin area values are then added up and averaged across the aggregation dataset (grid/plot/site) to form one return energy profile. In this profile the separation between ground and vegetation part of the profile is found as a local minimum, sought going upwards from the ground level within a pre-set number of bins. Finally the ground part of the profile is corrected for the vegetation-ground reflectance difference using either a fixed or dataset-adjusted ratio (discussed in Section 3.3).

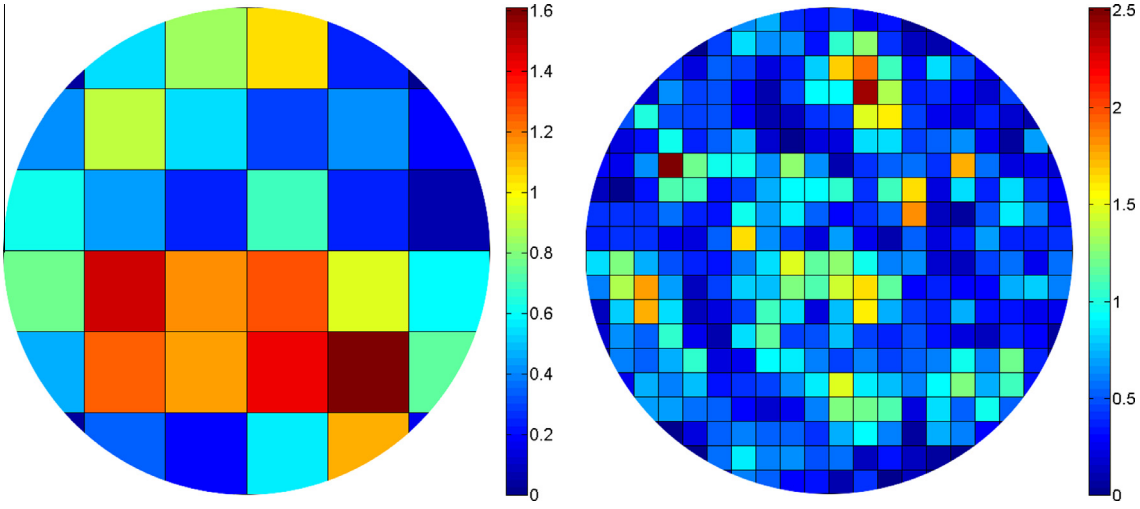

Fig. 4. Example of 15 m-radius plot-level (left) and 50 m-radius site-level (right) LAIe maps with 5 m grid cell size. The LAIe LiDAR values in each cell are weighted by the area covered by the cell in order to enable calculation of weighted average plot- and site-level LAIes. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

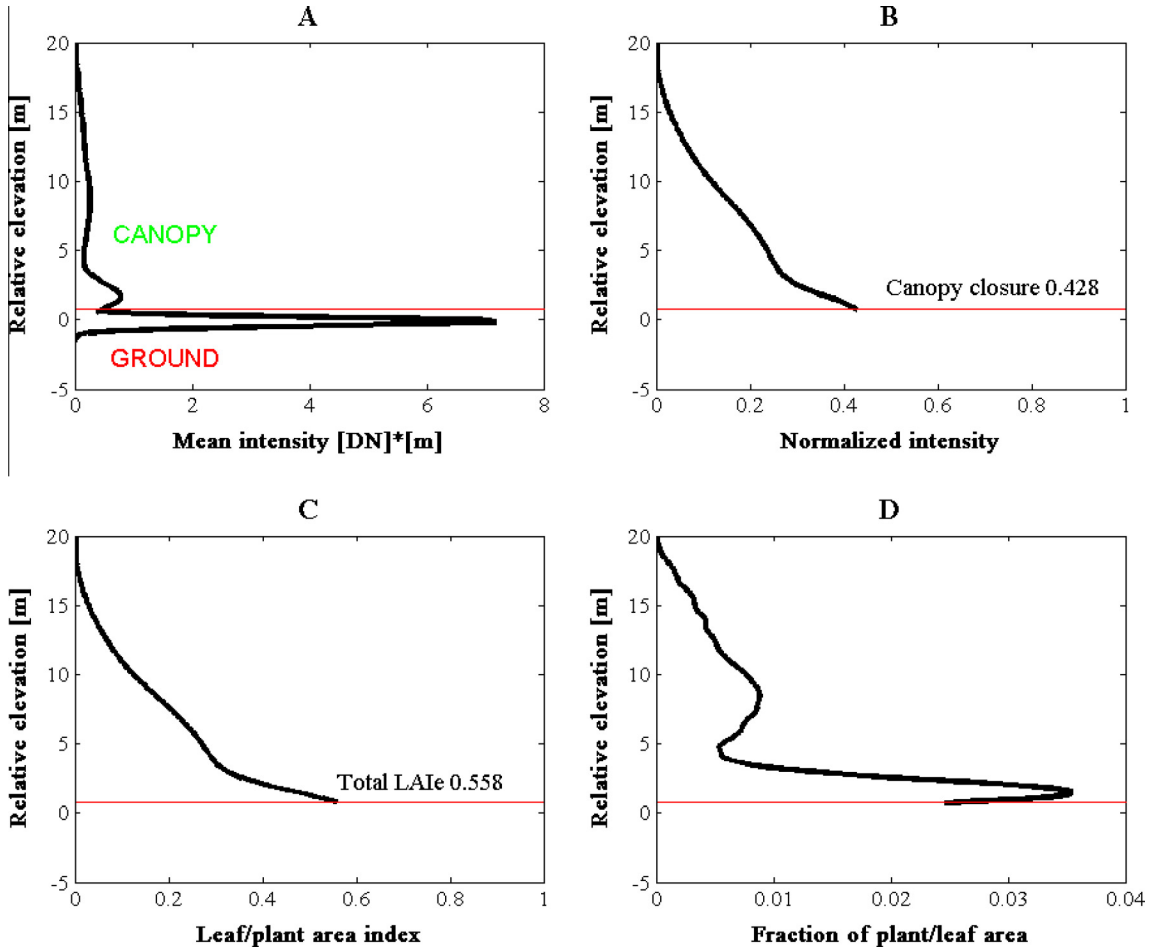

Fig. 5. Example of CHP processing stages for Site 10: (A) returned energy profile; (B) canopy closure profile; (C) cumulative leaf/plant area index profile; and (D) canopy height profile. Red line represents the beginning of ground return. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

3.2.3. Canopy closure profile

Canopy closure profile is computed by cumulatively adding up the bin values in the energy profile from the top of vegetation until the beginning of ground return and by normalizing it by the total cumulative energy including ground return (Fig. 5B).

3.2.4. Effective leaf area profile and index

To correct for the effect of occlusion the canopy closure profile is transformed to cumulative leaf (plant) area index profile (Fig. 5C). The last vegetation bin value is treated as the total LAIe for the dataset and used in comparisons with hemi-photos. The following transformation is applied:

$$LAIe = -\ln(1 - closure) \tag{1}$$

3.2.5. Canopy height profile

Finally, the canopy height profile is generated by turning the leaf area profile into an incremental distribution and normalizing it by the total value of leaf area index (Fig. 5D). In the case of grid-processed datasets the aggregation of cells is performed prior to normalization of the profile.

3.3. Dataset-adjusted reflectance ratio

Due to the difference in reflectance between the vegetation and the ground at the laser wavelengths (ground being approximately twice as reflective as vegetation), the value of bins corresponding to the ground in the energy profile had to be adjusted. Ni-Meister et al. (2010) proposed calculating the reflectance ratio using two neighbouring footprints, as the difference of their accumulated vegetation returns over the difference of their ground returns. This calculation was done with the assumption of constant reflectance ratio of neighbouring footprints. The method was applied to large-footprint LVIS data (~20 m). In the case of small-footprint LiDAR data such as that considered here, the problem in using this method is that the waveforms may not always have a ground return and a larger sample would be necessary to be representative of the area. Armston et al. (2013) proposed accounting for the reflectance difference between canopy (ρ_v) and ground (ρ_g) for small-footprint laser data using

$$\frac{\rho_v}{\rho_g} = -\frac{R_v}{R_g - J_o\rho_g} \tag{2}$$

where R_v is integrated vegetation return, R_g integrated ground return and J_o is the transmitted pulse energy corrected for transmission losses. The $J_o\rho_g$ term can be estimated as the mean integral of single-peak ground returns, assuming that ρ_g is constant and the mean converges to a normal distribution (Armston et al. (2013). The dataset-adjusted reflectance ratio was calculated at plot- and site-level as well as within each 5 m by 5 m grid. In cases where the ratio could not be calculated within certain grid cells due to small sample size, the corresponding mean plot-level reflectance ratio was used.

4. Results and discussion

The dataset-adjusted reflectance ratio, based on Armston et al. (2013), used in CHP methodology, was calculated at plot-, site- and grid-level. The mean value for all plots yielded a ratio of 0.60, with a minimum of 0.44, maximum of 0.74, standard deviation of 0.06, and standard error of the mean (SEM) of 0.008. The site-level reflectance ratio was calculated for the site-aggregated datasets with a center at the center of central plot and radius of 50 m. The mean of the site-level reflectance ratio of all sites also yielded 0.60 with minimum of 0.50, maximum of 0.67, standard deviation of 0.06, and SEM of 0.016. Fig. 6 presents the relationship between the integrated vegetation and ground returns R_v and R_g calculated from site-level data in 5 m grids. This plot shows a linear relationship between those integrals and confirms that the assumption of constant vegetation and ground reflectance relation is valid for this study site.

4.1. LAI validation

The LiDAR LAIe values at plot- and site-level, derived as one of the stages of the CHP procedure for the plot/site-aggregated and grid-processed datasets, were compared to the hemispherical photography-derived LAIe. The analysis was carried out for 54 plots and 11 sites. For clarity, Tables 1 and 2 summarize the acronyms of the methods used at plot and site-level, respectively.

4.1.1. Plot-level

Fig. 7 presents absolute differences between LiDAR and hemispherical photography-derived LAIe depending on the aggregation method and reflectance ratio, while Fig. 8 shows regression results between them. In general, all variants of the method provided estimates quite close to those of hemi-photos, however, there was a small tendency for very low hemispherical LAIes to be overestimated whereas for higher hemi-photo LAIes LiDAR showed some degree of underestimation. In plot-aggregated datasets, technique WF1 (with plot-adjusted reflectance ratio) provided lower estimates than WF2 (with a constant ratio), whose mean was closer to that of hemispherical photography. In the grid-processed datasets the patterns were very similar to those in plot-aggregated datasets. In discontinuous discrete canopies it is expected that grid-processed LAIe values should be higher compared to plot-aggregated estimates. This is due to the fact that as the LAIe is a logarithmic function of canopy penetration probability, and it is sensitive to cell heterogeneity; as cell area becomes smaller it contains more homogenous vegetation and data. Therefore when the average LAIe value of the whole plot area is calculated, the lack of canopy continuity (sparseness of the canopy) is taken into account, producing a more accurate plot average. Both grid-processed variants of the method provided higher LAIe estimates in comparison to plot-aggregated datasets, making the estimates even closer to those of hemi-photo values, with WF2g having a mean LAIe value almost the same as that of the hemispherical photography.

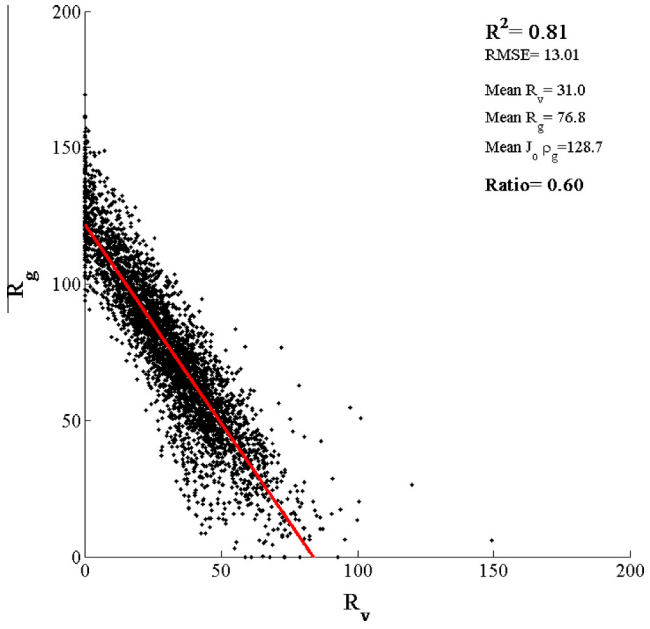

Fig. 6. Relationship between R_v and R_g for grid-processed site-level datasets.

Table 1
Summary of plot-level methods acronyms used with their description.

Method	Reflectance ratio	Aggregation
WF1	Adjusted	Aggregated (15 m radius cylinder)
WF2	Fixed	Aggregated (15 m radius cylinder)
WF1g	Adjusted	Gridded (5 m grid)
WF2g	Fixed	Gridded (5 m grid)

Table 2
Summary of plot-level methods acronyms used with their description.

Method	Reflectance ratio	Aggregation	Data used
WF1O	Adjusted	Aggregated	50 m radius cylinder
WF1A	Adjusted	Aggregated	Plot average
WF2O	Fixed	Aggregated	50 m radius cylinder
WF2A	Fixed	Aggregated	Plot average
WF1gO	Adjusted	Gridded	50 m radius cylinder
WF1gA	Adjusted	Gridded	Plot average
WF2gO	Fixed	Gridded	50 m radius cylinder
WF2gA	Fixed	Gridded	Plot average

Looking at Fig. 8 one can notice that the correlation with hemispherical photography at-plot level was rather moderate. Better results were achieved for WF1 with dataset-adjusted reflectance ratio, despite the fact that the mean of those LAIe values was in general lower than for WF2. Therefore, the dataset-adjusted reflectance ratio helped to improve the correlation with the hemi-photos. Furthermore, processing data in 5 m grid also improved the level of correlation, especially for WF2, showing that small-footprint data with aggregation area adjusted to the site heterogeneity may be advantageous for LAIe estimation in discontinuous canopy environments. The paired two-tailed *t*-tests carried out on the differences between each methods' and hemispherical photography LAIes showed that despite its lower correlation, only the mean of grid-processed WF2g was not significantly different (at the 5% significance level) from the mean of hemi-photo LAIes, providing a *p*-value of 0.67. The *t*-tests between LAIe estimates with different reflectance ratio (WF1 vs. WF2, WF1g vs.WF2g) as well as between plot-aggregated and grid-processed datasets (WF1 vs. WF1g, WF2 vs. WF2g) showed that the change of reflectance ratio or aggregation area causes the estimates to be significantly different.

There are several possible reasons for moderate (although still significant) correlation at plot-level, including the geo-coding uncertainties and consequent difficulties in locating exactly the same areas in LiDAR data as seen by the photography, as well as shading effects in LiDAR as a result of different incidence angles. Another potential error source is the fact that the LiDAR estimates are calculated from a limited and exactly known area (cylinders), whereas the hemi-photos do not have a specific footprint. Further, a complete agreement between the two methods of LAIe

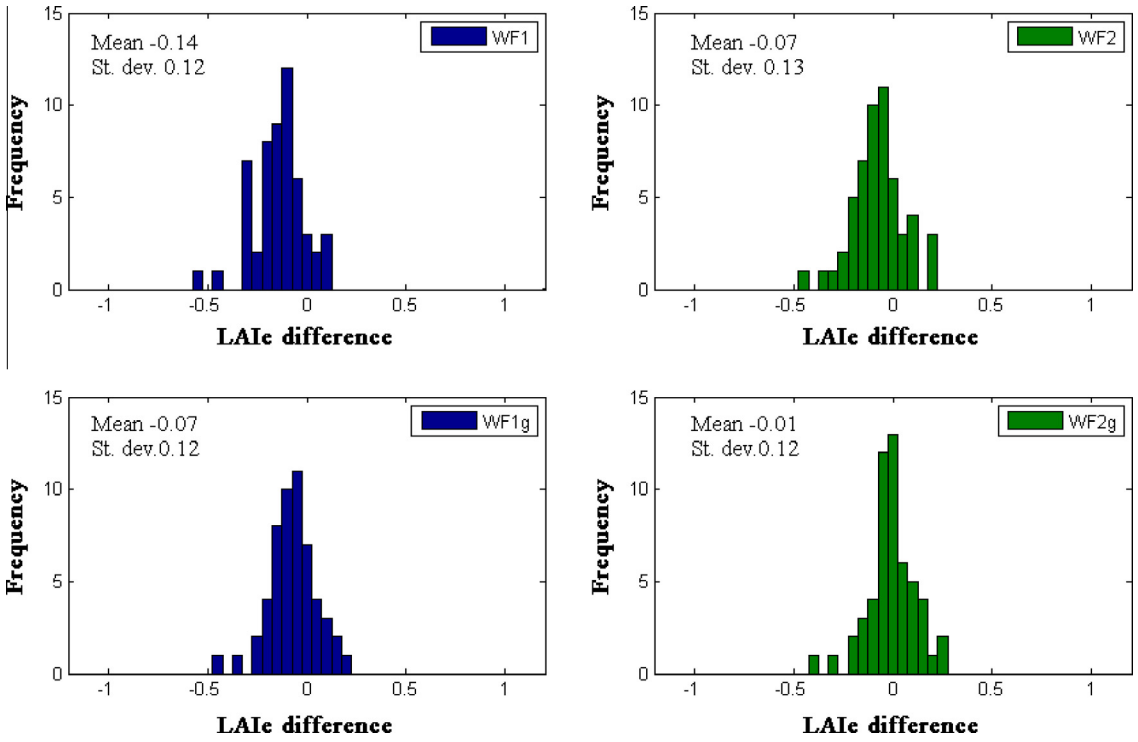

Fig. 7. Histograms of absolute differences between LiDAR and hemispherical photography LAIe estimates depending on the reflectance ratio and aggregation area used. The method names denoted with letter 'g' correspond to grid-processed datasets.

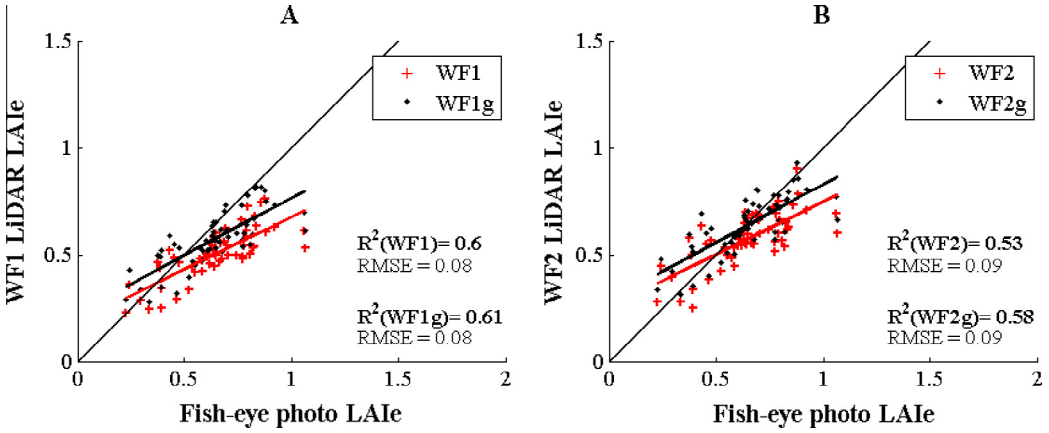

Fig. 8. Plot-level regression of LiDAR LAIes against hemi-photo LAIes. (A) WF1 and WF1g with dataset-adjusted reflectance ratio; (B) WF2 and WF2g constant reflectance ratio. The method names denoted with letter 'g' correspond to grid-processed dataset. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

estimation cannot be achieved due to different geometry of measurements: a cylinder which passes down through the canopy for LiDAR data and an upward looking cone for hemispherical photography (Solberg et al., 2006). Additionally, the hemispherical photography LAIe values themselves may not well represent the LAIe values in discontinuous canopies, and their acquisition conditions (the sun present in some frames) could have introduced estimation errors. Finally, the LAIe values over Gillenbah forest are rather low. This means that for higher LAIe values the method could potentially stabilise and provide better correlation with the hemi-photo estimates.

4.1.2. Site-level

Site-level LiDAR LAIe estimates were derived in two ways: (1) from the processing of LiDAR data within a 50 m radius from the center of the central plots ('Overall' – denoted with letter 'O' e.g. WF1O) and (2) from averaging the 15 m radius plot-level LiDAR LAIes within each site ('Average' – denoted with letter 'A' e.g. WF1A). The plots occupy a fraction of each site, as shown in Fig. 1 (inset), so are not necessarily representative of the entire site. Both options were performed on the site-aggregated as well as on grid-processed datasets. Since the hemispherical photography LAIe estimates were derived as mean values of plot-level estimates, LiDAR site-level LAIe estimates derived in the same way ('Average') provided better comparison (higher correlation values) than estimates derived from 50 m cylinders ('Overall'). Therefore only those results ('Average') will be discussed further, however the results for the 'Overall' method are also presented in figures and tables for reference as they should provide better estimates at site-level due to better coverage.

The patterns discussed at plot-level were repeated at the site-level still showing some level of underestimation of hemi-photo values by LiDAR. The correlation between the LiDAR data and hemi-photo estimates (Fig. 9), however, significantly improved at site-level and the RMSE errors dropped in comparison to plot-level values for both variants of the method. WF1A and WF1Ag yielded the highest correlation values, reaching R^2 of 0.83, yet again showing the superiority of using a dataset-adjusted reflectance ratio over a fixed ratio of 0.5, which in turn provided correlation of 0.59 and 0.76 for WF2A and WF2gA, respectively. The advantage of small-footprint data for LAIe estimation in discontinuous canopy environment was also confirmed at the site-level, especially in the 'Overall' approach to LAIe computation. As in the case of plot-level data, only estimates of grid-processed WF2g site-level LAIe turned out to be not significantly different to the hemi-photo values. Further, the tests between LAIe estimates with different reflectance ratio (WF1A vs. WF2A, WF1g vs. WF2g) as well as between plot-aggregated and grid-processed datasets (WF1A vs. WF1gA, WF2A vs. WF2gA) showed that the change of reflectance ratio and aggregation area causes the estimates to be significantly different.

4.2. CHP Validation

Canopy height profiles (CHP) generated from LiDAR data with a dataset-adjusted reflectance ratio (WF1) and a fixed reflectance ratio of 0.5 (WF2) were compared to the field biomass profiles modelled from field measured data. The comparison was conducted at plot- and site-level for plot/site-aggregated and grid-processed datasets using bin-wise ordinary least squares regression. Fig. 10 shows an example of profiles generated for each of the five

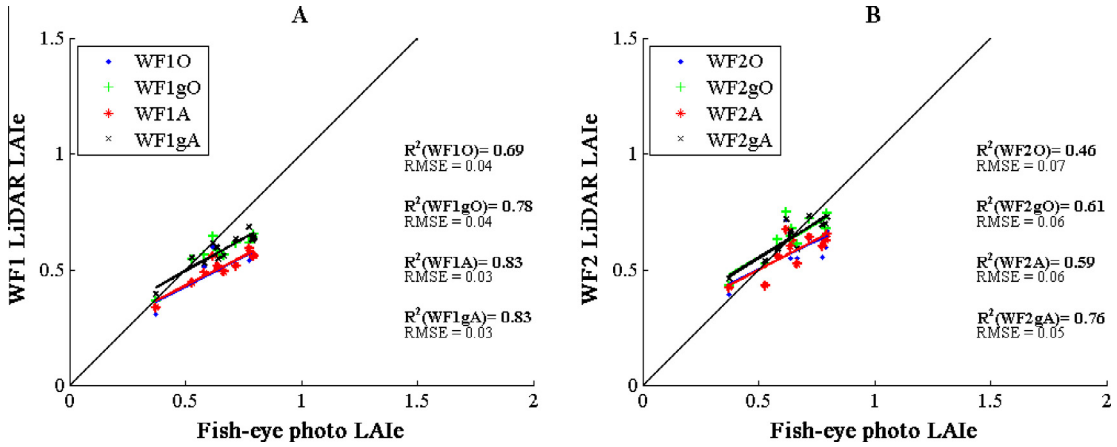

Fig. 9. Site-level regression of LiDAR LAIes against hemi-photo LAIe. (A) WF1 and WF1g with dataset-adjusted reflectance ratio; (B) WF2 and WF2g constant reflectance ratio; The method names denoted with letter 'g' correspond to grid-processed dataset whereas letter 'A' refers to the fact that the LAIe values were calculated as plot-level means. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

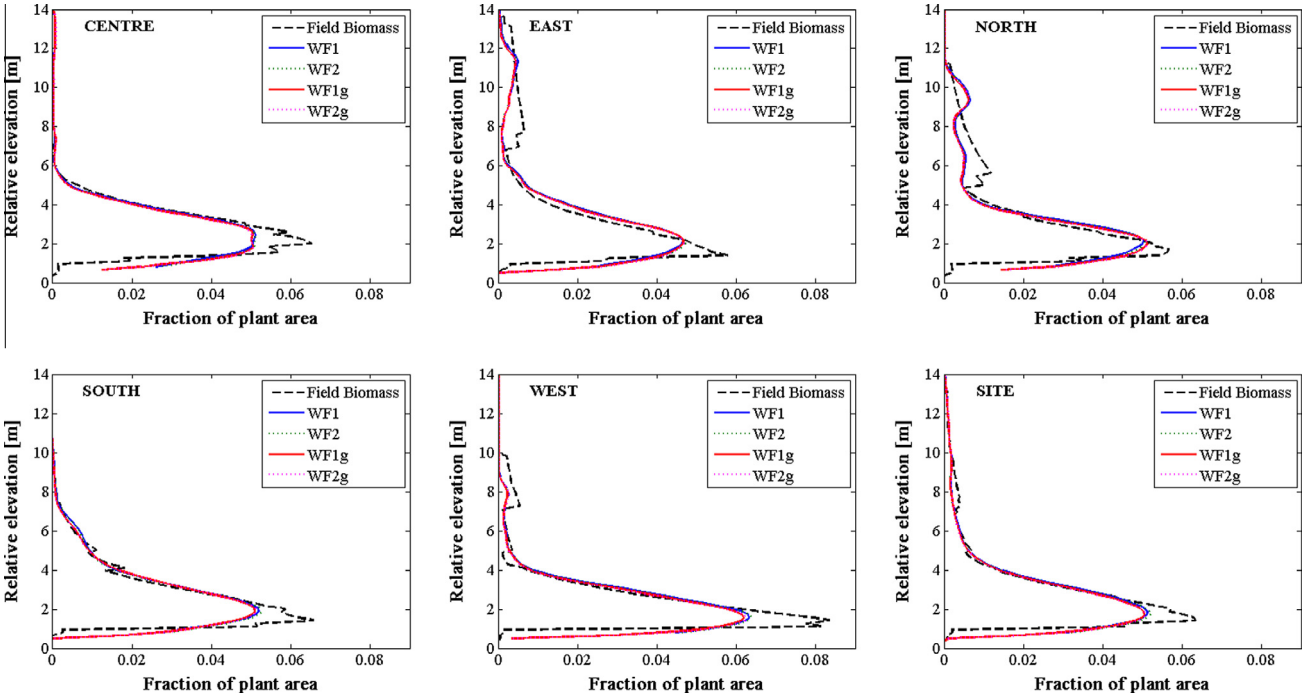

Fig. 10. Example of CHPs generated for Site 42 in comparison to field biomass profiles for the five 15 m-radius plots and the overall 50 m-radius site. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

plots, as well as site-level CHP in site 42 for plot/site-aggregated and grid-processed datasets with field profiles overlaid.

4.2.1. Plot-level

Fig. 11 shows the R^2 values of regression between LiDAR CHP and fieldwork-derived biomass profiles for all plots (grouped into the 11 sites), depending on the reflectance ratio used, for plot-aggregated and grid-processed datasets. The correlation between field and LiDAR data is excellent in the case of all plots in Sites 17, 20, 30 and 42. In those sites R^2 did not drop below 0.8 for almost all plots and plot-aggregated and grid-processed profiles provide very similar results. The only exceptions of correlation lower than 0.8 in those sites were: plot 17E with correlation of 0.72 in plot-aggregated and 0.78 in grid-processed datasets and plot 20S with correlation of 0.79 in grid processed datasets. The reason for lower correlation in those plots could have been the fact that neither of them had understory (in contrast to the remaining plots) and they consisted of the lowest number of trees, which would mean that the profiles would have been more sensitive to e.g. geolocation errors.

The comparison for Site 38 was also very good with high correlation values for most of the plots. The center and East plots had the lowest though still significant (above 0.6) correlation values. Similarly, with the exception of the north and west plots, site 24 would have also been one of those with the highest correlations. Site 10, in turn, yielded relatively consistent but somewhat lower correlation values ranging from 0.6 to 0.8. While it is hard to tell what could have been the reason for lower correlation in plot 38C and most of the plots of site 10, in the case of plots 38E, 24N and 24W it could have as well been the low number of trees and lack of understory that caused the drop in correlation values. This seems to be confirmed by the fact that the grid-processed dataset provided higher correlation values in those plots than plot-aggregated data, by accounting for discontinuity of the canopy cover. Furthermore, there were two tall shrubs and dead branches in plot 24W, and a fallen tree and branches in plot 38E that were not taken into account in the generation of field biomass profiles that could also have contributed to the lower correlation.

Sites 15, 23 and 67 were mostly quite inconsistent, with R^2 values varying between close to zero and over 0.9. In those sites, correlation values from plot-aggregated datasets differed the most in comparison to grid-processed datasets, with the latter yielding usually higher values. The lowest correlations were achieved for plots 15C and 15N, where there were tree trunks and branches on the ground, and plots 23C, 23S, and 67E, with relatively small number of trees, in plot-aggregated datasets. With the exception of plot 15C, the corresponding R^2 values in grid-processed datasets were moderate and mostly above 0.6.

Finally, the R^2 values in site 99 were very low – between 0.2 and 0.6. There are several reasons for that. Site 99 had a relatively high percentage of very tall *E. microcarpa* trees with diameters up to 8–10 times larger than those of typical *C. glaucophylla* trees. This site also had the highest number of dead trees (23%) in comparison to other sites. Plots 99N and 99W, which yielded the lowest correlation values, had the lowest number of trees (<25) and about 30% of trees without measured CBHs. Plots 99E, 99S and 99W also had the highest percentages of dead trees (26–27%). Since the dead trees in Gillenbah forest were mostly just single trunks, they would have been relatively easily missed by the near-vertical LiDAR beams, whereas by being included in the field profiles, they could have further contributed to the differences between field and LiDAR data.

Summarizing, the sites with the highest values of R^2 in their plots (sites: 17, 20, 30, 42) are the sites with the densest vegetation and highest numbers of trees measured (over 210 trees). The only exception is site 23 which also has dense vegetation (248 trees in total) but somewhat lower correlation values. The reason for that may again be the fact that this site has over 50% of crown base heights not measured and tree heights estimated from allometric equations making field biomass profiles less reliable. The plots with the low correlation values typically have a low number (<25) of trees measured (15C, 15N, 23C, 38E, 67E) or if the number of trees is slightly higher (25–35) those plots usually include some

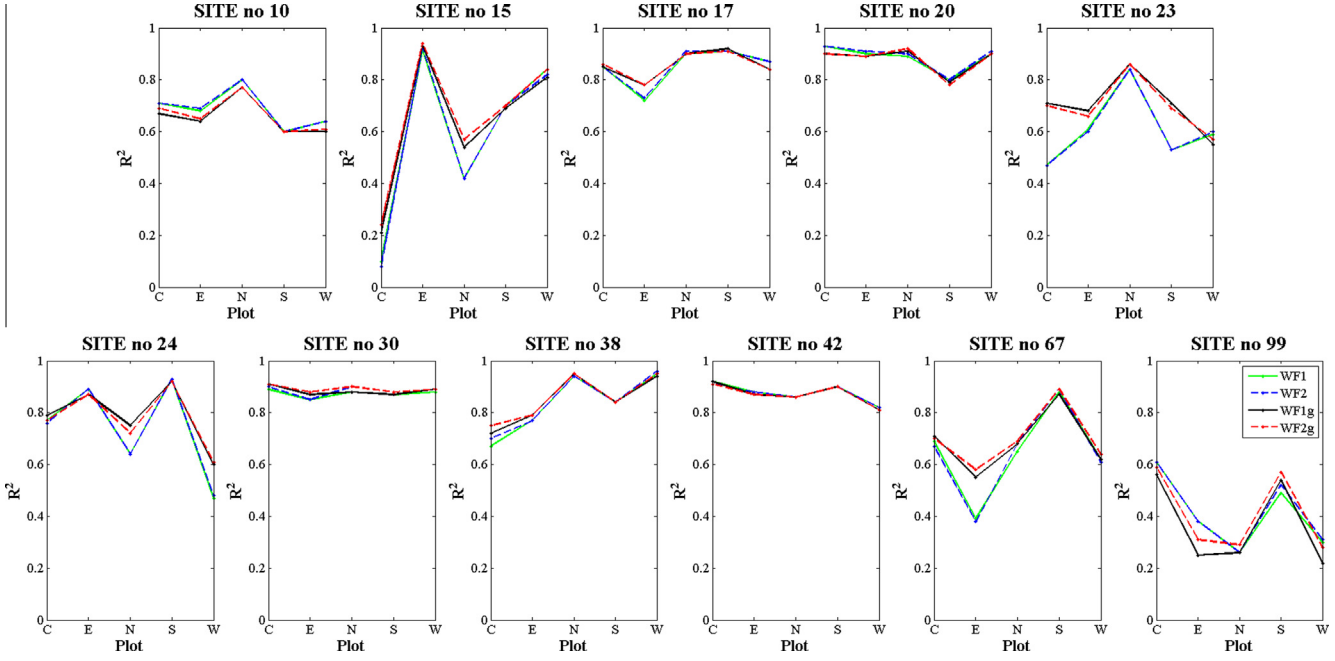

Fig. 11. R^2 values of CHP regression against field biomass profiles at plot-level in each site for plot-aggregated and grid–processed datasets, with different reflectance ratios used. The values on the *x* axis correspond to a plot within each site (C – Central, E – East, N – North, S – South, W – West). (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

Table 3
Mean plot-level and site-level R^2 and RMSE of CHP regression against field biomass profiles.

Method	Reflectance ratio	Plot-level 55 plots		Site-level 11 sites	
		RMSE	R^2	RMSE	R^2
Plot/site-aggregated	WF1	0.005	**0.73**	0.004	**0.85**
	WF2	0.005	**0.73**	0.004	**0.86**
Grid-processed	WF1	0.004	**0.74**	0.003	**0.85**
	WF2	0.004	**0.75**	0.003	**0.86**

percentage of *E. microcarpa* trees (10C, 24W). This reiterates the importance of accurate geo-location in the case of sparse canopies.

Table 3 summarizes the R^2 and RMSE for each of the variants of comparisons tested at plot-level. The mean R^2 of 0.73 for 55 plots was achieved regardless of reflectance ratio used in plot-aggregated datasets. When grid-processed CHPs were compared the average correlation slightly increased. Using a constant reflectance ratio produced a further small improvement (0.75). The averaged RMSEs were constant for all the comparisons with marginally higher values for plot-aggregated datasets.

The *t*-tests carried out between LiDAR and field biomass profiles at plot-level showed that all but one CHP plot, regardless of reflectance ratio used or aggregation area, were not statistically significantly different from field biomass profiles. The only exception was the grid-processed profile of plot 20S generated with 0.5 ratio that turned out to be significant different from field profile, while its corresponding profile generated with a dataset-adjusted reflectance ratio was on the border of significance. The comparison between different reflectance ratios proved that the CHPs generated with dataset-adjusted reflectance ratio were not significantly different (with *p*-values close to 1) to those obtained with fixed ratio of 0.5, regardless of the aggregation area. When plot-aggregated and grid-processed datasets were confronted, 27% of CHPs with dataset-adjusted reflectance ratio and 27% of CHPs with fixed ratio proved to be significantly different at 5% of significance level. A further 7–8% were on the borderline of significance. Therefore, about a third of grid-processed profiles showed some statistical differences in comparison to their corresponding plot-aggregated profiles, suggesting that in discontinuous canopy environments it may be worth using small-footprint data with an aggregation area size adjusted to capture the heterogeneity of vegetation in the scene.

4.2.2. Site-level

The site-aggregated and grid-processed CHPs derived from the 50 m radius sites were compared to the field profiles. Fig. 12 presents correlation for each site for site-aggregated and grid-processed datasets, with different reflectance ratios used, compared to field biomass profiles, while Table 3 shows mean R^2 and RMSE values of 11 sites. The patterns of correlation were in general the same as at plot-level, however, with a noticeable increase in R^2. The mean correlation at site-level was higher than the corresponding mean plot-level R^2 by 0.11–0.13. In the site-aggregated and grid-processed datasets, the choice of site-adjusted or constant reflectance ratio of 0.5 had no significant impact on the correlation of the field and LiDAR data.

Both the site-aggregated and grid-processed datasets yielded almost identical site-level correlation values with field biomass profiles. Slightly higher mean R^2 values of 0.86 were provided by WF2 with a constant reflectance ratio, however WF1 with a dataset-adjusted ratio yielded an only marginally worse value of 0.85. The RMSEs also marginally improved in comparison to plot-level. The *t*-tests between site-aggregated LiDAR CHPs and field profiles confirmed the plot-level results that LiDAR profiles, regardless of

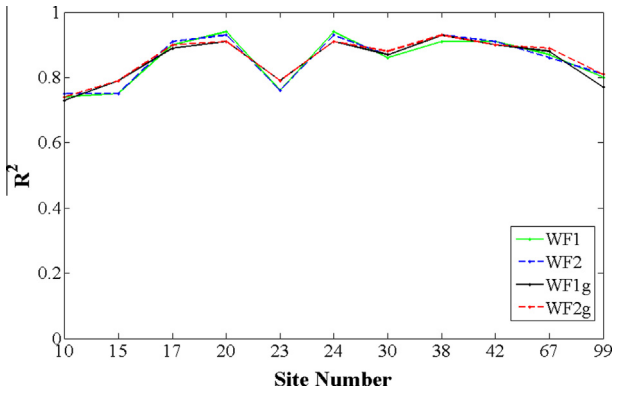

Fig. 12. R^2 values of CHP regression against field biomass profiles at site-level for grid-processed and site-aggregated datasets, with different reflectance ratios used. (For interpretation of the references to color in this figure legend, the reader is referred to the web version of this article.)

reflectance ratio used or aggregation area, were not significantly different than field biomass profiles. The *t*-tests carried out between CHPs generated with different reflectance ratios once again proved no significant difference (*p*-values close to 1) between the LiDAR profiles from WF1 and WF2 at site-level, regardless of aggregation area. Comparison between site-aggregated and grid-processed datasets, similarly to plot-level data, showed that one third of the profiles generated in the 5 m grid were significantly different to the corresponding site-aggregated datasets.

5. Conclusions

This study has presented a validation of the Canopy Height Profile methodology (based on Harding et al. 2001) for leaf area index and vertical foliage retrieval, using small-footprint (0.2 m) full-waveform airborne LiDAR data in a discontinuous canopy cover environment, at plot (30 m diameter) and site-levels (100 m diameter). The methodology has been enhanced to include a dataset-adjusted reflectance ratio according to Armston et al. (2013) and tested against a fixed ratio of 0.5. This study has also looked at the influence of the aggregation area on LAIe and CHP retrievals in discontinuous canopy cover environments by carrying out calculations of LAIe and CHPs on grid-processed (5 m) and plot/site-aggregated (15 m/50 m radius) datasets. The analyses have shown that this methodology is relevant and suitable for its purpose.

The LAIe extracted as an intermediate step of the methodology was compared to hemispherical-photography estimates. The analysis has shown that the method produced LAIe values which were close to those of hemispherical photography with some degree of underestimation. The correlation values were somewhat moderate (however still significant) at the plot-level owing most likely to geo-coding uncertainties and a lack of GPS measurements of the hemispherical photography locations. At site-level the correlation improved significantly. Due to the fact that hemi-photo LAIe estimates at site-level were derived as a mean value of plot-level LAIes, the comparison to LiDAR site-level estimates was always better when the site-level LiDAR estimates were derived in the same way, rather than from a cylinder with 50 m radius. The highest R^2 value of 0.83 at site-level was reached by WF1 with dataset-adjusted reflectance ratio. Using a dataset-adjusted reflectance ratio as opposed to a fixed ratio of 0.5 did considerably improve the correlation with hemispherical photography values both at plot- and site-level, proving the importance of accurate estimation of the vegetation-ground reflectance ratio in LAIe

estimation and the value of the Armston et al. (2013) methodology for this application.

The influence of the aggregation area on LAIe estimates of discontinuous canopies was assessed by performing the calculation of LAIe in 5 m by 5 m grids versus datasets aggregated to plot- (15 m radius) and site-level (50 m radius). Gridding improved the correlation between LiDAR and hemi-photo-derived LAIe estimates both at plot- and site-level especially when fixed reflectance ratio was used. The study has, therefore, shown that in the case of discontinuous canopies it is important to consider heterogeneity of vegetation by choosing the aggregation size of LiDAR data that enables depiction of its sparseness. This may suggest that for such environments small-footprint LiDAR data (here with a footprint diameter of 0.2 m) may be particularly suitable as they offer the possibility of selecting a cylinder radius or grid size for aggregation that ensures homogeneity of the area covered by it. Choosing an aggregation area that is larger and does not depict the heterogeneity of vegetation or using LiDAR data with a footprint size that does not allow for it will most likely lead to underestimation of the LAIe due to the logarithmic transformation involved in the LAIe calculation. For areas with DTM variations and significant slopes, the use of small-footprint LiDAR data would be even more advantageous as it would help to avoid the influence of the terrain on the shape of the received waveforms and in consequence inaccuracies in the estimates. As a final remark, it is worth pointing out that due to its better spatial coverage and less effort in terms of collecting data LiDAR-based estimation of LAIe should provide better estimates than hemispherical photography (Morsdorf et al., 2006; Richardson et al., 2009).

Comparison of LiDAR CHPs to field biomass profiles generated based on field measurements made at the time of LiDAR acquisition, has proven that this methodology is also appropriate for description of the vertical distribution of vegetation. The correlation between field and LiDAR profiles was very high and reached a maximum mean R^2 of 0.75 at plot-level and 0.86 at site-level when 55 plots and the corresponding 11 sites were considered. All LiDAR plot-level profiles but one and all site-level profiles were also found to be not statistically significantly different (at 5% significance level) to field biomass profiles. Unlike the results of the LAIe determination, due to relative character of the CHP, using dataset-adjusted or constant reflectance ratio had almost no impact on the correlation of LiDAR with field biomass profiles (the fixed ratio gave marginally better correlation by 0.01). Furthermore, the profiles generated with different ratios were found to be not significantly different between each other (with p-values close to 1), regardless of the aggregation type and both at plot- and site-level. In the case of the grid-processed datasets, the correlation between LiDAR and field biomass data was usually marginally higher (by 0.01–0.02) than that of plot/site-aggregated datasets (both at plot- and site-level). Moreover, around one third of the grid-processed LiDAR canopy height profiles were found to be statistically significantly different from the corresponding plot/site-aggregated profiles. This suggests that not only for LAIe but also for foliage vertical profiles, using small-footprint airborne data may be advantageous in discontinuous canopy environments. Further experiments are, however, required in order to verify the performance and the use of the methodology at denser forest areas and at areas with a wider range of LAIe values.

Acknowledgements

This work was funded by Engineering and Physical Sciences Research Council (Grant number: EP/P505682/1, F4011613), School of Mathematical and Physical Sciences and School of Systems Engineering of University of Reading. Field campaign data was funded by Australian Research Council projects LE0560930, DP09845861 and FS100100040. Waveform processing was funded by the National Centre for Earth Observation. The authors would like to thank two anonymous reviewers for their extremely useful comments, which helped to make this article more interesting. Thanks also to A. McGrath and W. Lieff at Airborne Research Australia for undertaking the LiDAR flights and to all the SMAPEx-3 experiment participants: M. Bürgin, A. Colliander, Y. Gao, S. Hasan, D. Jinyang, A. Joseph, S. Kim, L. McKee, A. Monerris, P. O'Neill, J. Ouellette, R. Panciera, C. Rudiger, G. Satalino, C. Vittucci, F. Winston, X. Wu, and H. Yardley.

References

Aber, J.D., 1979. Method for estimating foliage-height profiles in broad-leaved forests. J. Ecol. 67 (1), 35–40.

Armston, J., Disney, M., Lewis, P., Scarth, P., Phinn, S., Lucas, R., Bunting, P., Goodwin, N., 2013. Direct retrieval of canopy gap probability using airborne waveform lidar. Remote Sens. Environ. 134, 24–38.

Black, T.A., Chen, J.-M., Lee, X., Sagar, R.M., 1991. Characteristics of shortwave and longwave irradiances under a Douglas-fir forest stand. Can. J. For. Res. 21 (7), 1020–1028.

Breda, N.J.J., 2003. Ground-based measurements of leaf area index: a review of methods, instruments and current controversies. J. Exp. Bot. 54 (392), 2403–2417.

Burrows, W.H., Hoffman, M.B., Compton, J.F., Back, P.V., 2001. Allometric relationships and community biomass stocks in White Cypress Pine (*Callitris glaucophylla*) and associated eucalpyts of the Carnarvon area, central Queensland. National Carbon Accounting System Technical Report No. 33. The Australian Greenhouse Office.

Calders, K., Lewis, P., Disney, M., Verbesselt, J., Armston, J., Herold, M., 2012. Effects of clumping on modelling LiDAR waveforms in forest canopies. IEEE International Geoscience and Remote Sensing Symposium, Munich, Germany, pp. 3391–3394.

Chen, J.M., Black, T.A., 1992. Defining leaf area index for non-flat leaves. Plant, Cell Environ. 15 (4), 421–429.

Chen, J.M., Cihlar, J., 1996. Retrieving leaf area index of boreal conifer forests using Landsat TM images. Remote Sens. Environ. 55 (2), 153–162.

Chen, J.M., Black, T.A., Adams, R.S., 1991. Evaluation of hemispherical photography for determining plant area index and geometry of a forest stand. Agric. For. Meteorol. 56 (1–2), 129–143.

Chen, J.M., Rich, P.M., Gower, S.T., Norman, J.M., Plummer, S., 1997. Leaf area index of boreal forests: theory, techniques, and measurements. J. Geophys. Res.: Atmos. 102 (D24), 29429–29443.

Chen, X.T., Disney, M.I., Lewis, P., Armston, J., Han, J.T., Li, J.C., 2014. Sensitivity of direct canopy gap fraction retrieval from airborne waveform lidar to topography and survey characteristics. Remote Sens. Environ. 143, 15–25.

Fieber, K.D., Davenport, I.J., Ferryman, J.M., Gurney, R.J., Walker, J.P., Hacker, J.M., 2013a. Analysis of full-waveform LiDAR data for classification of an orange orchard scene. ISPRS J. Photogram. Remote Sens. 82, 63–82.

Fieber, K.D., Davenport, I.J., Tanase, M.A., Ferryman, J.M., Gurney, R.J., Walker, J.P., Hacker, J.M., 2013b. Preliminary leaf area index estimates from airborne small footprint full-waveform LiDAR data. IEEE International Geoscience and Remote Sensing Symposium, Melbourne, Australia, pp. 3379–3382.

Fieber, K.D., Davenport, I.J., Tanase, M.A., Ferryman, J.M., Gurney, R.J., Walker, J.P., Hacker, J.M., 2014. Effective LAI and CHP of a single tree from small-footprint full-waveform LiDAR. IEEE Geosci. Remote Sens. Lett. 11 (9), 1–5.

Hamilton, S.D., Brodie, G., O'dwyer, C., 2005. Allometric relationships for estimating biomass in grey box (*Eucalyptus microcarpa*). Aust. Forest. 68 (4), 267–273.

Harding, D.J., Lefsky, M.A., Parker, G.G., Blair, J.B., 2001. Laser altimeter canopy height profiles: methods and validation for closed-canopy, broadleaf forests. Remote Sens. Environ. 76 (3), 283–297.

Hopkinson, C., Lovell, J., Chasmer, L., Jupp, D., Kljun, N., Van Gorsel, E., 2013. Integrating terrestrial and airborne lidar to calibrate a 3D canopy model of effective leaf area index. Remote Sens. Environ. 136, 301–314.

Hosoi, F., Nakai, Y., Omasa, K., 2010. Estimation and error analysis of woody canopy leaf area density profiles using 3-D airborne and ground-based scanning lidar remote-sensing techniques. IEEE Trans. Geosci. Remote Sens. 48 (5), 2215–2223.

Jensen, J.L.R., Humes, K.S., Vierling, L.A., Hudak, A.T., 2008. Discrete return lidar-based prediction of leaf area index in two conifer forests. Remote Sens. Environ. 112 (10), 3947–3957.

Jonckheere, I., Fleck, S., Nackaerts, K., Muys, B., Coppin, P., Weiss, M., Baret, F., 2004. Review of methods for in situ leaf area index determination: Part I. Theories, sensors and hemispherical photography. Agric. For. Meteorol. 121 (1–2), 19–35.

Koetz, B., Morsdorf, F., Sun, G., Ranson, K.J., Itten, K., Allgower, B., 2006. Inversion of a lidar waveform model for forest biophysical parameter estimation. IEEE Geosci. Remote Sens. Lett. 3 (1), 49–53.

Lang, A.R.G., Mcmurtrie, R.E., Benson, M.L., 1991. Validity of surface area indices of Pinus radiata estimated from transmittance of the sun's beam. Agric. For. Meteorol. 57 (1–3), 157–170.

Lefsky, M.A., 1997. Application of Lidar Remote Sensing to the Estimation of Forest Canopy and Stand Structure. University of Virginia.
Lefsky, M.A., Cohen, W.B., Acker, S.A., Parker, G.G., Spies, T.A., Harding, D., 1999. Lidar remote sensing of the canopy structure and biophysical properties of douglas-fir western hemlock forests. Remote Sens. Environ. 70 (3), 339–361.
Lefsky, M.A., Harding, D., Keller, M., Cohen, W.B., Carabajal, C., Del Bom Espirito-Santo, F., Hunter, M., De Oliveira Jr., R., 2005a. Estimates of forest canopy height and aboveground biomass using ICESat. Geophys. Res. Lett. 32.
Lefsky, M.A., Hudak, A.T., Cohen, W.B., Acker, S.A., 2005b. Geographic variability in lidar predictions of forest stand structure in the Pacific Northwest. Remote Sens. Environ. 95 (4), 532–548.
Levy, P.E., Jarvis, P.G., 1999. Direct and indirect measurements of LAI in millet and fallow vegetation in HAPEX-Sahel. Agric. For. Meteorol. 97 (3), 199–212.
Lindberg, E., Olofsson, K., Holmgren, J., Olsson, H., 2012. Estimation of 3D vegetation structure from waveform and discrete return airborne laser scanning data. Remote Sens. Environ. 118, 151–161.
Lovell, J.L., Jupp, D.L.B., Culvenor, D.S., Coops, N.C., 2003. Using airborne and ground-based ranging lidar to measure canopy structure in Australian forests. Can. J. Remote Sens. 29 (5), 607–622.
Macarthur, R.H., Horn, H.S., 1969. Foliage profile by vertical measurements. Ecology 50 (5), 802–804.
Macfarlane, C., Hoffman, M., Eamus, D., Kerp, N., Higginson, S., Mcmurtrie, R., Adams, M., 2007. Estimation of leaf area index in eucalypt forest using digital photography. Agric. For. Meteorol. 143 (3–4), 176–188.
Mallet, C., Bretar, F., 2009. Full-waveform topographic lidar: State-of-the-art. ISPRS J. Photogram. Remote Sens. 64 (1), 1–16.
Miller, M.E., Lefsky, M., Pang, Y., 2011. Optimization of Geoscience Laser Altimeter System waveform metrics to support vegetation measurements. Remote Sens. Environ. 115 (2), 298–305.
Monerris, A., Walker, J.P., Panciera, R., Jackson, T., Tanase, M., Gray, D., Ryu, D., 2011. The Third Soil Moisture Active Passive Experiment. Workplan.
Morsdorf, F., Kötz, B., Meier, E., Itten, K.I., Allgöwer, B., 2006. Estimation of LAI and fractional cover from small footprint airborne laser scanning data based on gap fraction. Remote Sens. Environ. 104 (1), 50–61.
Morsdorf, F., Nichol, C., Malthus, T., Woodhouse, I.H., 2009. Assessing forest structural and physiological information content of multi-spectral LiDAR waveforms by radiative transfer modelling. Remote Sens. Environ. 113 (10), 2152–2163.
Ni-Meister, W., Lee, S., Strahler, A.H., Woodcock, C.E., Schaaf, C., Yao, T., Ranson, K.J., Sun, G., Blair, J.B., 2010. Assessing general relationships between aboveground biomass and vegetation structure parameters for improved carbon estimate from lidar remote sensing. J. Geophys. Res.: Biogeosci. 115 (G2), G00E11.
Panciera, R., Walker, J.P., Jackson, T.J., Gray, D.A., Tanase, M.A., Ryu, D., Monerris, A., Yardley, H., Rüdiger, C., Wu, N., Gao, Y., Hacker, J.M., 2013. The Soil Moisture Active Passive Experiments (SMAPEx): toward soil moisture retrieval from the SMAP mission. IEEE Trans. Geosci. Remote Sens. (99), 1–18
Riaño, D., Valladares, F., Condés, S., Chuvieco, E., 2004. Estimation of leaf area index and covered ground from airborne laser scanner (Lidar) in two contrasting forests. Agric. For. Meteorol. 124 (3–4), 269–275.
Richardson, J.J., Moskal, L.M., Kim, S.-H., 2009. Modeling approaches to estimate effective leaf area index from aerial discrete-return LiDAR. Agric. For. Meteorol. 149 (6–7), 1152–1160.
Solberg, S., Næsset, E., Hanssen, K.H., Christiansen, E., 2006. Mapping defoliation during a severe insect attack on Scots pine using airborne laser scanning. Remote Sens. Environ. 102 (3–4), 364–376.
Solberg, S., Brunner, A., Hanssen, K.H., Lange, H., Næsset, E., Rautiainen, M., Stenberg, P., 2009. Mapping LAI in a Norway spruce forest using airborne laser scanning. Remote Sens. Environ. 113 (11), 2317–2327.
Sun, G., Ranson, K.J., 2000. Modeling lidar returns from forest canopies. IEEE Trans. Geosci. Remote Sens. 38 (6), 2617–2626.
Tang, H., Dubayah, R., Swatantran, A., Hofton, M., Sheldon, S., Clark, D.B., Blair, B., 2012. Retrieval of vertical LAI profiles over tropical rain forests using waveform lidar at La Selva, Costa Rica. Remote Sens. Environ. 124, 242–250.
Weiss, M., Baret, F., Smith, G.J., Jonckheere, I., Coppin, P., 2004. Review of methods for in situ leaf area index (LAI) determination: Part II. Estimation of LAI, errors and sampling. Agric. For. Meteorol. 121 (1–2), 37–53.
White, M.A., Asner, G.P., Nemani, R.R., Privette, J.L., Running, S.W., 2000. Measuring fractional cover and leaf area index in arid ecosystems: digital camera, radiation transmittance, and laser altimetry methods. Remote Sens. Environ. 74 (1), 45–57.
Wilson, J.W., 1959. Analysis of the spatial distribution of foliage by two-dimensional point Quadrats. New Phytol. 58 (1), 92–99.
Wilson, J.W., 1965. Stand structure and light penetration. I. Analysis by Point Quadrats. J. Appl. Ecol. 2 (2), 383–390.
Yao, T., Yang, X., Zhao, F., Wang, Z., Zhang, Q., Jupp, D., Lovell, J., Culvenor, D., Newnham, G., Ni-Meister, W., Schaaf, C., Woodcock, C., Wang, J., Li, X., Strahler, A., 2011. Measuring forest structure and biomass in New England forest stands using Echidna ground-based lidar. Remote Sens. Environ. 115 (11), 2965–2974.
Zhao, F., Yang, X., Schull, M.A., Román-Colón, M.O., Yao, T., Wang, Z., Zhang, Q., Jupp, D.L.B., Lovell, J.L., Culvenor, D.S., Newnham, G.J., Richardson, A.D., Ni-Meister, W., Schaaf, C.L., Woodcock, C.E., Strahler, A.H., 2011. Measuring effective leaf area index, foliage profile, and stand height in New England forest stands using a full-waveform ground-based lidar. Remote Sens. Environ. 115 (11), 2954–2964.
Zhao, F., Strahler, A.H., Schaaf, C.L., Yao, T., Yang, X., Wang, Z., Schull, M.A., Román, M.O., Woodcock, C.E., Olofsson, P., Ni-Meister, W., Jupp, D.L.B., Lovell, J.L., Culvenor, D.S., Newnham, G.J., 2012. Measuring gap fraction, element clumping index and LAI in Sierra Forest stands using a full-waveform ground-based lidar. Remote Sens. Environ. 125, 73–79.
Zheng, G., Moskal, L.M., 2012. Computational-Geometry-Based Retrieval of Effective Leaf Area Index Using Terrestrial Laser Scanning. IEEE Trans. Geosci. Remote Sens. 50 (10), 3958–3969.